图书情报工作
第65卷 第16期 2021年8月

面向政府开放数据的分面检索研究*

■ 林鑫[1,2] 龙存钰[1] 罗宇[1]

[1] 华中师范大学信息管理学院 武汉 430079 [2] 湖北省数据治理与智能决策研究中心 武汉 430079

摘 要：[目的/意义]构建面向分面检索的分面体系和政府开放数据分面检索模型，以优化用户体验。[方法/过程]立足政府开放数据特征进行政府开放数据分面体系构建，在此基础上，构建政府开放数据分面检索模型，并进行原型系统的实现以验证模型的可行性和效果。[结果/结论]通过原型系统实现证实所构建的政府开放数据分面体系能够有效帮助用户进行检索结果的探索和筛选；所提出的分面检索系统模型具有良好的用户体验，操作便捷且能够规避信息过载问题。

关键词：政府开放数据 分面检索 分面体系

分类号：G251

DOI：10.13266/j.issn.0252-3116.2021.16.014

分面检索大大提升了用户的信息检索效率，同时也改善了用户的使用体验[1-2]，在政府开放数据检索实践中得到了广泛应用。面向分面检索的分面体系及其展现控制是影响用户使用体验的核心要素，然而从现有政府开放数据的网站调研来看，存在着分面设置与展现控制上与用户认知习惯不一致的问题，导致用户检索体验不佳。分面设置上，问题表现有三：其一，部分政府网站开放数据内容相关的分面存在焦点重叠的混乱情况，导致分面体系可用性较差，如深圳市政府数据开放平台同时存在焦点重合度较高的领域、主题和行业三个分面；其二，部分政府网站的分面体系不完善，导致不能全面揭示政府开放数据特征；其三，存在分面的常用焦点设置不合理的情况，导致分面价值不能有效发挥，或者分面的焦点过多导致信息过载等问题[3]，如浙江省数据开放平台提供的数据格式分面上，其焦点包括“XSL”“XML”“JSON”等，但显然用户不太可能因为数据格式问题，而放弃获取某一开放数据集。展现控制上，现有政府开放数据网站缺乏根据检索结果动态调整的机制，易因为分面或焦点过多引发信息过载。

除此之外，当前的政府开放数据分散分布于各层级、区域的政府开放数据平台上，缺乏资源的集成发现。基于此，本文拟从用户需求出发，面向政府开放数据发现构建分面体系，进而构建面向多源数据的政府开放数据集成检索系统模型，以支持政府开放数据的集成发现。

1 相关研究与实践

在设计分面检索模型时，除了设计常规检索系统的所有功能模块外，还需要进行分面体系构建、分面体系应用策略设计和交互界面设计。其中，与本研究密切相关的是分面体系构建和分面体系应用策略设计，下面将围绕分面体系构建方法、分面应用策略以及面向政府开放数据的分面检索实践进展这三个方面进行综述。

1.1 分面体系构建方法

分面体系构建方法研究主要关注分面化改造、分面体系的完善以及分面的自动化构建。在分面化改造上，胡昌平和林鑫提出基于主题词表的分面化改造来构建科技文献内容分面[4]，高文飞和赵新力提出了基于《综合电子政务主题词表》的分面化改造构想[5]，贾君枝等基于FAST主题词分面应用原理和理念构造了适用于电子政务主题词表的7个基本分面和13个扩展分面[6]。在分面体系的完善上，相关研究关注的核心问题是如何更全面、深入地揭示资源的内容特征和质量特征。围绕这一问题，部分研究提出将用户生成

* 本文系国家社会科学基金青年项目“社会网络中基于用户认知结构的知识标注研究”（项目编号：17CTQ024）研究成果之一。

作者简介：林鑫（ORCID：0000-0003-0318-8160），副教授，博士；龙存钰（ORCID：0000-0001-5096-7461），硕士研究生，通讯作者，E-mail：longcunyu@mails.ccnu.edu.cn；罗宇（ORCID：0000-0002-1496-7066），本科生。

收稿日期：2021-02-07 修回日期：2021-05-27 本文起止页码：130-137 本文责任编辑：杜杏叶

内容、评分、热度等新的属性特征加入分面体系中[7-9]；还有研究则在已有分面的基础上进行完善，如政治人物检索中根据政治家孩子的毕业院校设置新的分面等[10]。尽管为保障分面与焦点选取的合理性、可理解性，专家法是分面体系构建的主流方法[11-12]，但分面体系半自动化与自动化的构建策略也已经开始探索。分面体系半自动化是指在人工设置分面体系框架的基础上，通过信息抽取进行焦点集的自动构建，如F. Abel通过实体识别技术进行火灾发生地、相关组织、人员三个分面的焦点集构建[13]；分面体系自动化则以实现分面体系框架设置与焦点集构建的全部自动化为目标，如窦志成等提出了融合检索结果与海量语料资源的分面体系自动挖掘策略[14]。

1.2 分面体系应用策略

为充分发挥分面体系的导航作用，同时规避其带来的信息过载问题，在进行分面应用时，需要设计分面体系排序策略、展现控制策略与交互机制。

1.2.1 分面体系排序策略

分面较少时，通常采用的策略是人工指定固定顺序。该策略可以在保障分面排序整体合理的前提下，为用户提供稳定的预期排序策略，提高了用户的分面利用效率，是实践中最为流行的策略。但也有研究围绕分面动态排序策略开展研究，以适应分面数量较多的情形，如E. Kharlamov等综合考虑了应用分面后获得的结果数量多寡、能否得到独特结果、浏览路径的丰富性3个因素设计排序策略[15]；孔盛球等提出了基于保持率的相关性判断方法[16]。在分面内的焦点排序上，常用的动态排序策略是按照相关结果的数量进行降序，常用的固定顺序策略包括人工指定、字顺排序、时间顺序、空间顺序等[17]。

1.2.2 展现控制策略

合理的展现控制策略有助于改善信息过载问题，随着分面体系的不断丰富与完善，其重要性得到了进一步加强。对分面的展现与否，E. Kharlamov等认为应收起不太重要的焦点与分面[18]，李兵提出应该根据用户查询意图来确定是否展现分面[19]。在层级化分面体系的展现形式上，胡昌平等提出应根据检索结果分布动态进行分面体系层数调整，以避免分面体系的深度过深[20]。

1.2.3 分面交互机制

该方面主要涉及同一分面内及不同分面间多个焦点间的运算逻辑。当前，研究与实践中较为通行的处理方案是同一个分面内的焦点间进行逻辑“或”运算，不同分面的焦点间逻辑“与”运算[21]。但也有研究指出，用户选择同一个分面内的多个焦点时，其需求也可能是进行逻辑“与”运算[22]。

1.3 相关实践进展

目前我国上线了130余个政府数据开放平台，其中绝大部分平台都提供了分面检索服务，下面以浙江省、贵州省、山东省和贵阳市较为典型的四个政府开放数据平台为例对当前政府开放数据的分面检索实践状况进行梳理。浙江省政府开放数据平台包含近万个数据集，其分面检索服务提供了领域、格式以及数据来源单位3个分面，无论用户检索结果分布如何，各分面的焦点均全部呈现；贵州省政府开放数据平台提供1 400多个数据集，其分面检索服务提供了主题、场景、部门、类型、格式以及开放属性等6个分面，前三个分面的焦点默认不展现，后三个分面直接呈现所有焦点；山东省政府开放数据平台提供近千个数据集，其分面检索服务提供领域、格式、开放类型以及省直部门4个分面，各分面的焦点均全部直接显示，但只有省直部门分面展示各焦点相关的数据集数量；贵阳市政府开放数据平台提供近3 000个数据集，其分面检索服务提供领域、主题、行业、服务、市级部门、区县列表、数据格式等7个分面。交互机制上，上述四个平台均不支持同一分面内选择多个焦点，不同分面之间的焦点均是逻辑“与”的关系。

总体来说，国内外围绕分面体系构建方法与分面应用策略开展了多方面研究，其成果对政府开放数据分面检索研究与实践具有参考意义。但是，针对政府开放数据的分面检索理论研究仍然较为缺乏，而实践探索虽取得了初步进展，但仍存在较为突出的问题：分面体系构建多来自于开放数据的外部特征，对数据内容特征揭示不足，部分分面对用户检索结果的探索与过滤价值不大，易引发信息过载；交互机制不够灵活，不支持分面内同时选择多个焦点及分面体系展现的动态控制，影响用户使用体验。为解决这些问题，本文拟先立足用户的分面应用需求进行分面体系重构，进而构建面向用户的政府开放数据分面检索模型，以优化政府开放数据分面检索服务的使用体验。

2 政府开放数据分面体系构建

当前，政府开放数据主要通过国家及地方政府的开放数据平台进行发布，每个数据集都只有唯一来源。因此，在当前的信息环境下，尽管政府开放数据具有多方面的特征，如数据格式、数据规模、发布时间、数据主

题、发布部门等,但用户只需要利用数据主题、类型(如名单、公示、统计)、发布机构、时间范围几个特征即可实现检索结果的探索与过滤,获得唯一来源的数据集。用户在进行开放数据检索时,会优先考虑能够辅助检索的数据特征。一般来说,用户会结合开放数据的内容特征和外部特征来进行筛选。在内容特征上,用户会优先考虑开放数据的主题和类型,通常通过前者来快速判断该检索对象的内容是否符合自身的检索需求,并通过后者定位所需的某一种类的开放数据。在外部特征上,用户会优先考虑开放数据的发布机构和更新时间,即用户通过数据发布机构来判断开放数据的级别等信息,以及通过数据更新时间考虑时效性。在分面体系构建中,鉴于数据发布机构的数量众多,为减少信息过载,需要将其进一步细化为行政区划级别、行政区划名称、发布机构类型 3 个分面;数据集所涉及数据的时间信息(如采集时间、统计时间范围等)难以全部获取,但基本可以通过数据的更新时间进行推测,因此将数据更新时间纳入分面体系。根据以上认识,政府开放数据分面体系应包含开放数据主题、开放数据类型、行政区划级别、行政区划名称、发布机构类型以及数据更新时间 6 个分面,如图 1 所示:

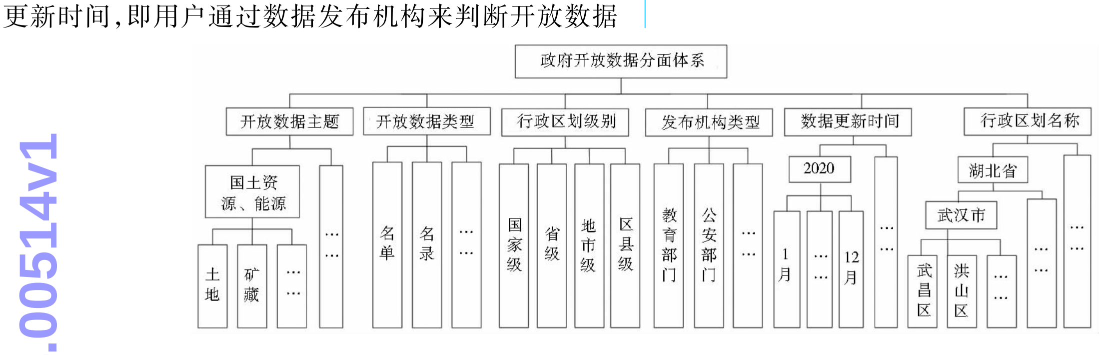


图 1　政府开放数据分面体系(部分)

2.1　开放数据主题分面

不同的政府开放数据网站提供不同的数据主题分面,导致目前的开放数据主题分面较为混乱,没有形成一个统一的标准,而《政务信息资源目录体系第 4 部分:政务信息资源分类》(GB/T 21063.4－2007)[23] 提出的主题类目体系包括 21 个一级类目和 133 个二级类目,较为全面地涵盖了政务信息资源的主题,在实践中得到了广泛应用,研究采用该类目体系作为政府开放数据的主题分面依据。从设置的类目来看,该体系的部分一级类目过于泛化,不能很好地进行主题划分,而且下属的二级类目差异较大,如一级类目"国土资源、能源"下设有"土地""矿藏""水资源""海洋""煤炭""石油""燃料、燃气""电力"以及"综合类"等 9 个二级类目,同时,"综合类"这个二级类目也出现在其他一级类目下。因此,如果只用一级类目作为每个分面的焦点,粒度过粗,难以帮助用户准确定位所需资源;如果只用二级类目会造成每个主题下的"综合类"混淆,且焦点数量过多,易造成信息过载。基于此,拟将 GB/T 21063.4－2007 中的一级和二级类目均纳入主题分面,形成一个包含亚面的层级分面体系。

2.2　开放数据类型分面

通常一个数据集只涵盖了其所涉及主题的一个或几个方面,或者属于不同的公文种类,据此可以将数据集划分为不同的类型,以表明政府开放数据集所包含数据的性质、用途,如名单、指标、明细、公示、通知。鉴于当前政府开放数据缺乏类型的系统划分,研究采用了归纳法进行开放数据类型的分面焦点设置。首先采集了包括贵州、广东在内的 9 个省市政府开放数据集名称,作为构建分面焦点词集合的基础数据;在此基础上,采用人工分析与机器辅助相结合的方式对全部 21 036条数据进行分析,提取其中体现数据类型的特征词;最后,人工对提取结果进行规范化和归并、去重处理,形成数据类型分面的焦点词集合,共包含清单等 46 个。

2.3　行政区划级别、行政区划名称和发布机构类型分面

随着时间的发展和政府开放数据的不断完善,政府机构作为数据发布主体参与政府开放数据建设的积极性将会提高,发布机构的数量将日益增加。若进行跨区域、跨层级甚至是全国性的政府开放数据发现平台构建时,直接将发布机构作为焦点,将会带来较为严重的信息过载和利用不便问题。为解决这一问题,研究拟设置行政区划级别、名称和发布机构类型 3 个分面,从而既可以通过单个焦点实现多个机构的复合选择,也可以通过跨分面焦点的组合实现发布机构的精确定位。

(1)行政区划级别分面。我国现行的行政区划共

有四级:省级行政区、地市级行政区、区县级行政区和乡镇级行政区。实践中,乡镇级的开放数据数量较少,而国家级的开放数据数量较多且开放程度较高,因此,参考我国的四级行政区划,将开放数据的区划级别分面的焦点设置为国家级、省级、地市级和区县级。

(2)行政区划名称分面。区划名称指的是开放数据发布机构所属的具体行政区划,如湖北省、郑州市、杭州市余杭区等。该分面是包含省、地市、区县的三层结构,可以通过区划间的归属关系进行层层定位。

(3)发布机构类型分面。我国行政体制下,尽管不同层级的地方政府之间,及地方政府与中央政府之间在机构设置上具有较强的一致性,如中央设有教育部,地方设有教育厅、教育局,但也有所差异,如地方政府普遍设置的政务服务数据管理局在国家层级上就没有相应的部门。因此,考虑国家机关设置,以包括上海、山东等在内的 8 个省级和地级市政府开放数据平台上的数据为基础,进行了发布机关类型分面焦点词的归纳,最终得到的焦点词包括 122 个,如教育部门、公安部门、网信部门等。

2.4 数据更新时间分面

政府开放数据集具有时效性,用户在进行开放数据集选择时,会考虑到该数据集的更新时间,以获取所需时间范围内的信息。理论上来讲,更新时间是一个连续的概念,而且随着时间的推移,取值也会发生变化。实际应用过程中,结合其他分面,精确到开放数据的更新年月就能够满足用户的结果过滤需求。因此,更新时间分面设置年份和月份两个层级,以帮助用户便捷地选择具体年月。

3 政府开放数据分面检索系统模型

当前,我国政府开放数据的发布由各级政府自主完成,开放数据的分布非常分散,因此为实现政府开放数据分面检索,首先需要进行元数据的采集、整合,并在此基础上依据政府开放数据分面体系进行标注,以实现数据集与分面体系的关联;进而设计分面体系的排序、展现控制、交互机制等应用策略。基于以上过程,政府开放数据分面检索系统架构见图 2,包含政府开放数据元数据采集、整合模块,自动标引模块、分面应用策略模块、检索模块以及交互界面模块。政府开放数据元数据采集、整合模块为自动标引模块提供数据支持,也是分面体系设计和分面应用策略模块的基础;自动标引模块将开放数据整合模块和分面体系模块关联起来;开放数据检索模块和分面应用策略模块将计算机与用户连接起来。其中,交互界面设计可以参考文献[24－25]的研究成果,而检索模块则与普通检索系统类似,不再展开讨论,下文将重点介绍开放数据元数据采集、整合与标引模块以及分面应用策略模块。

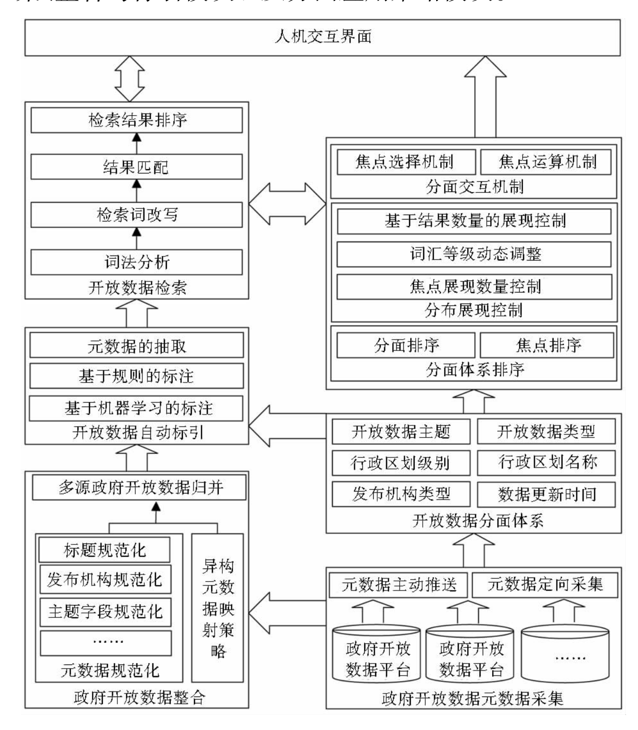


图 2 政府开放数据分面检索系统模型

3.1 政府开放数据元数据采集

为实现政府开放数据的集成发现,首先需要进行元数据的采集,此处的元数据除了包括各级政府发布的政府开放数据元数据外,还包括各数据集的 URL 信息。元数据采集既可以通过与各政府开放数据平台进行协作,即由平台定期将新增或发生更新的数据集的元数据信息推送过来,也可以利用网络信息采集工具的定向采集方式主动去各政府数据开放平台进行元数据获取。采用定向采集方式时,除了需要采集新增的数据集外,还需要对已采集的数据集信息进行监测,及时发现更新数据。

3.2 政府开放数据整合

当前的开放数据元数据采集均以官方平台为信息源,且各数据集均具有唯一性,因此,政府开放数据整合的主要工作是元数据的规范化及异构元数据的映射策略设计。元数据的规范化是指对采集来的元数据进行数据转换、信息增补等以提升数据的规范性。其中,需要特别注意的是标题与发布机构的规范化、主题元数据规范化。标题规范化是指需要对未明确体现开放数据所针对地域信息的数据集标题进行处理(如“白云区统计月报”,难以确定是贵阳市的白云区还是广州

市的白云区),使其明确包含所属地市、区县信息,以避免歧义、提升标题的可理解性。发布机构的规范化与其相似,需要将不完备的机构名称补充完整。主题元数据规范化是指部分政府开放数据平台并未采用 GB/T 21063.4－2007 进行主题著录,而是采用了数据领域或行业词表,这就需要基于两个词表间的映射关系将其转换为 GB/T 21063.4－2007 中对应的主题词。异构元数据映射策略设计是指建立不同开放数据平台的元数据项映射规则,以解决政府开放数据平台采用的元数据标准不同,或者对于同一项元数据的命名不同而无法整合的问题。

3.3　政府开放数据集标注

在实现数据整合的基础上,还需要按照前文构建的政府开放数据分面体系进行标注,从而为资源检索及分面体系应用奠定数据基础。数据集标注主要通过现有元数据抽取、基于规则的标注和基于机器学习的标注进行实现。其中,元数据抽取指的是开放数据集已有元数据已经包含了相应的信息,将其提取出来即可实现标注,其主要适用于开放数据主题、数据更新时间的标注;基于规则的标注是指通过抽取规则从开放数据集的相关信息中提取出基础信息,进而通过映射规则建立数据集与相应分面的焦点的关联,该方法主要应用于行政区划级别、行政区划名称、发布机构类型的标注;基于机器学习的标注是指利用机器学习方法进行模型训练,以实现开放数据主题、开放数据类型的标注。

3.4　分面应用策略模块

为提高用户使用分面的效率,减少分面展现带来的信息过载问题,需要进行分面体系排序策略、展现控制策略和交互机制策略设计。

(1)分面体系排序。鉴于政府开放数据分面体系中仅包含六个分面,数量不多且均为常用分面,因此研究采用固定排序的策略,分面顺序依次为开放数据主题、开放数据类型、行政区划级别、发布机构类型、行政区划名称和数据更新时间。不同的分面下,焦点排序策略也有所不同,行政区划级采用固定顺序分面,按级别从高到底排列;数据更新时间分面下,年份按照倒序排列,月份按照从 12 月到 1 月进行排列;其余分面按照焦点覆盖的开放数据数量从多到少排列。

(2)分面体系展现控制。当检索结果较少时,用户无需通过分面体系对结果进行探索或过滤,此时不展现分面体系。而当分面体系展现时,通过调整词汇等级、控制默认展现的焦点数量进行展现控制。调整词汇等级是指当开放数据的主题词包含下位词,且包含该主题词的开放数据集中于一个或少数几个下位词时,则将该主题词从分面体系中剔除,同时将下位词上移,简化分面体系的层级。控制默认展现的焦点数量是指当检索结果对应的焦点数量过多时,依据焦点间的排序结果,默认只展现前面几个焦点,同时支持用户自主查看全部焦点。

(3)分面交互机制。为提高分面交互的灵活性,同一个分面内或不同分面间允许焦点的多选,同一分面内的焦点间设置为逻辑“或”的关系,不同分面的焦点间设置为逻辑“与”的关系。

4　政府开放数据分面检索系统原型实现

为验证政府开放数据分面检索模型的可行性和效果,研究采集了广东省、贵州省等 6 个省级政府开放数据平台以及滨州市、深圳市、杭州市等 24 个地市级政府开放数据平台开放的数据集作为基础数据(采集时间为 2020 年 12 月 25 日至 30 日),进行了该原型系统的实现。原型系统采用 B/S(Browser/Sever,浏览器/服务器)架构,处理器是 Intel(R) Core(TM) m3 －8100Y CPU @ 1.10GHz 1.61GHz,运行内存 4GB;软件的操作系统平台为 Windows 10,基于 x64 的处理器。

4.1　原型系统构建过程

得益于各省市政府开放数据平台均注重数据的标准化,因此各数据集内部的规范性均较强。然而,各省市之间的开放数据存在较为严重的数据异构情况,包括元数据体系差异较大,同一个元数据项命名不同、取值规范不同。以广东省和杭州市政府开放数据平台为例,前者包含后者所没有的“关键字”元数据项,前者将“发布机构”元数据项称为“数据提供方”,后者则称为“发布部门”;前者的发布日期信息精确到日,后者则精确到秒。同时,各省市元数据中的标题、发布机构和主题著录也存在不规范的现象。基于此,研究在数据整合环节,首先结合各开放数据平台的元数据体系建立了原型系统的元数据体系;继而,对无法精确匹配的元数据项名称,建立了各省市政府开放数据平台与原型系统间的映射关系,以解决元数据名称异构问题;最后,对取值不规范或异构的元数据进行处理,保持各省市开放数据的一致性,实现了数据的整合。

数据标注环节中,原型系统重点关注的是前文所构建的分面体系在数据集中的体现,因此需要标注数据集的主题、类型、行政区划级别与名称、发布机构类型以及数据更新时间。其中,数据主题与数据更新时间通过元数据抽取实现;行政区划名称标注采用规则

林鑫，龙存钰，罗宇. 面向政府开放数据的分面检索研究[J]. 图书情报工作,2021,65(16):130－137.

从数据标题中抽取；行政区划级别通过区划名称与级别的映射关系获取；发布机构类型标注中，首先通过发布机构元数据项获取相应信息，从中提取机构类型特征词，进而根据映射词表进行标注；数据类型的标注同样采用了规则法，如果数据类型下的焦点词出现在数据集的标题中，则以该焦点词作为其数据类型，否则标注为综合类。

分面应用策略设计上，分面及焦点的排序、分面交互机制参照前文模型实施；展现控制上，检索结果少于10条时（一页最多展现10条结果），展现分面的价值不大，且可能对用户造成干扰，故而不再展现分面；默认状态下，每个分面最多展现5个焦点或亚面，以避免信息过载问题；当某一亚面下的焦点只有一个或者两个（除行政区划名称外），则不展现该亚面，而是将焦点等级提升一级，直接展现出来；分面下只有一个焦点时，隐藏该分面。

4.2 原型系统效果

总体来说，政府开放数据分面检索原型系统的用户体验良好，系统提供的分面可帮助用户便捷地进行检索结果探索与过滤，同时受展现控制策略的影响，也基本不会带来信息过载问题。研究以搜寻各个地方的新冠肺炎救治医院名单数据为例进行展示，以直观反映原型系统的效果。首先以“新型冠状病毒”为查询式进行检索，结果如图3所示。此时，检索结果数量较多，需要分面体系的辅助才能便捷地进行结果筛选。从左侧展示的分面体系看，展示的主题（包括医疗保健、政务综合类等焦点）、类型（包括名单、公示等分面）、发布机构（卫生健康部门、当地政府等焦点）等分面可以帮助用户快速进行结果筛选；同时，受益于展现控制策略，默认展现的分面和焦点数量较少，体系较为扁平化，不会带来信息过载问题，检索体验良好。鉴于需求是救治医院的名单，因此选择开放数据主题中的“医疗保健”和数据类型中的“名单”作为条件进行过滤。如图4所示，检索结果从97条精简为12条，数据相关性有了非常明显的改善，前5条结果中4条是相关结果。通过这一交互过程可以看出，本文所构建的政府开放数据分面体系和分面检索系统模型的可行性和良好效果，能够为用户进行政府开放数据搜寻提供有效支持，大幅提升其搜寻效率和交互体验。



图3　原型系统初始检索结果界面



图 4　选择焦点“医疗保健”和“名单”后的检索结果

5　结语

在当前的政府开放数据分面检索实践中,由于分面体系设置和应用策略设计不够合理,用户使用体验不佳。为了解决这一问题,本文在构建政府开放数据分面体系基础上,设计了政府开放数据分面检索系统模型,并以 6 个省级政府开放数据平台和 24 个政府地市级开放数据平台的数据为基础进行了原型系统的实现。结果显示,研究所构建的政府开放数据分面体系能够帮助用户快速筛选和定位所需开放数据,分面检索系统模型中提出的政府开放数据整合策略、分面体系应用策略都非常有效,可以在为用户探索和筛选检索结果提供帮助的同时,减轻用户的认知负担,也不会带来信息过载的问题。

参考文献:

[1] KULES B, CAPRA R. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface[J]. Journal of the American Society for Information Science and Technology, 2012, 63(1): 114 – 138.

[2] MAHDI M N, AHMAD A R, ISMAIL R. Paradigm extension of faceted search techniques a review[J]. Journal of telecommunication, electronic and computer engineering, 2017, 9(2/7): 149 – 153.

[3] 林鑫,吴茜. 文献资源发现系统分面检索功能比较研究[J]. 数字图书馆论坛,2019,4(9):16 – 23.

[4] 胡昌平,林鑫. 科技文献检索中基于主题词表分面化改造的分面构建[J]. 情报学报, 2015, 34(8):875 – 884.

[5] 高文飞,赵新力. 我国政务信息资源主题分类研究[J]. 情报科学,2008,26(9):1340 – 1343.

[6] 贾君枝,武晓宇. 基于 FAST 的综合电子政务主题词表分面式改造[J]. 图书情报工作,2014,58(8):105 – 110.

[7] 陈果,肖璐,孙建军. 面向网络社区的分面式导航体系构建——以丁香园心血管论坛为例[J]. 情报理论与实践,2017,40(10):112 – 116.

[8] 王军,周妍. 电商平台的分面元数据研究[J]. 情报学报, 2016, 35(3):317 – 325.

[9] VANDIC D, DAM J W V, FRASINCAR F. Faceted product search powered by the semantic Web[J]. Decision support systems, 2012, 53(3):425 – 437.

[10] ARENAS M, GRAU B C, KHARLAMOV E, et al. SemFacet: semantic faceted search over yago[C]// International conference on world wide web. Seoul: ACM, 2014:123 – 126.

[11] 谢泽宇,施国良,杨汉钰,等. 国际淡水争端领域分面本体构建与应用[J]. 情报杂志,2018,37(11):192 – 196.

[12] KONG W. Extending faceted search to the open-domain web[J]. ACM SIGIR forum, 2016,50(1): 90 – 91.

[13] ABEL F, HAUFF C, HOUBEN G J, et al. Twitcident: fighting fire with information from social web streams[C]// International conference on world wide web. Lyon: ACM, 2012:305 – 308.

[14] 窦志成,江政宝,李谨秀,等. 基于词项图分析的查询分面挖掘方法[J]. 计算机学报, 2017, 40(3):556 – 569.

[15] KHARLAMOV E, GIACOMELLI L, SHERKHONOV E, et al. SemFacet: making hard faceted search easier[C]// ACM on conference on information and knowledge management. Singapore: ACM, 2017:2475 – 2478.

[16] 孔盛球,冯钧,杜丙帅. 水利元数据动态分面搜索引擎的设计与实现[J]. 计算机技术与发展, 2017, 27(10):151 – 155.

[17] 邱明辉. 信息查询系统的分面导航设计研究[J]. 现代情报，2018,38(10):78－84.

[18] KHARLAMOV E, GIACOMELLI L, SHERKHONOV E, et al. SemFacet: making hard faceted search easier[C]// ACM on conference on information and knowledge management. Singapore: ACM, 2017:2475－2478.

[19] 李兵. 基于查询意图识别的自适应图书分面检索研究[J]. 图书馆学研究, 2017(15):57－64.

[20] 胡昌平,林鑫. 科技文献检索中基于主题词表分面化改造的分面构建[J]. 情报学报, 2015, 34(8):875－884.

[21] DEUSCHEL T, GREPPMEIER C, HUMM B G, et al. Semantically faceted navigation with topic pies[C]// International conference on semantic systems. Leipzig: ACM,2014:132－139.

[22] ARENAS M, GRAU B C, EVGENY E, et al. Towards semantic faceted search [C]// International conference on World Wide Web. Seoul: ACM, 2014:219－220.

[23] 全国信息技术标准化技术委员会. 政务信息资源目录体系 第4部分:政务信息资源分类: GB/T 21063.4－2007[S]. 北京: 中国标准出版社,2007:9.

[24] CAPRA R G, MARCHIONINI G. The relation browser tool for faceted exploratory search[C]//Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries. Pittsburgh: ACM, 2008: 420－420.

[25] DANIEL T. Faceted search[J]. Synthesis lectures on information concepts, retrieval, and services, 2009,1(1): 1－80.

作者贡献说明：

林鑫:负责论文选题,论文撰写与修订;

龙存钰:负责模型和原型的实现,初稿撰写;

罗宇:参与原型实现和文献资料搜集

Research on Faceted Search of Government Open Data

Lin Xin[1,2] Long Cunyu[1] Luo Yu[1]

[1] School of Information Management of Central Normal University, Wuhan 430079

[2] Hubei Data Governance and Intelligent Decision Research Center, Wuhan 430079

Abstract: **[Purpose/significance]** A government open data faceted system and a government open data faceted search model for faceted search are constructed to optimize user experience. **[Method/process]** Based on the characteristics of government open data, a faceted system was constructed. On this basis, a faceted search model of government open data was constructed, and the prototype system was implemented to evaluate the feasibility and effectiveness of the model. **[Result/conclusion]** The realization of the prototype system proves that the constructed government open data facet system can effectively help users to explore and screen search results. The proposed faceted search system model has a good user experience, which is convenient to use and can avoid the problem of information overload.

Keywords: government open data faceted search faceted system